

PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2003-248687**

(43)Date of publication of application : **05.09.2003**

(51)Int.Cl.

G06F 17/30

(21)Application number : **2002-046233**

(71)Applicant : **NIPPON YUNISHISU KK**

(22)Date of filing : **22.02.2002**

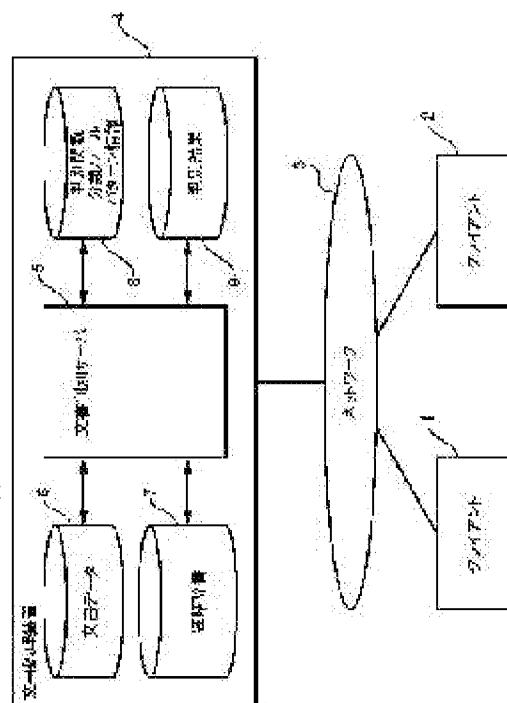
(72)Inventor : **MATSUDA YOSHIO
HAYASHIDA HIDEO**

(54) INFORMATION PROCESSOR AND ITS METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To solve a problem in mechanically determining a document and showing only a result, wherein a determination accuracy and the reliability of the determination result are obscure so as to have a possibility that a categorized opinion of a customer is sent around among corresponding divisions according to the determination dependent on an individual and is left aside for a long time.

SOLUTION: This information processor selects a copula of high appearance frequency from a document 10 (S20), creates a determination function 8a for determining a category of a document from the selected copula (S30), determines the category of the document 10 based on the created determined function 8a (S40), creates a classification rule of another result and classifies the determination result (S50-S60), calculates a validity of the determination result for every classification result (S70), and stores the determination function, the classification rule, and the validity in a memory as document determination information. This processor determines the category of a document 6 in a same type of the document 10 based on the document determination information and outputs the determination result 9 including the validity (S110-S140).



LEGAL STATUS

[Date of request for examination] 22.02.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3677006

[Date of registration] 13.05.2005

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2003-248687

(P2003-248687A)

(43)公開日 平成15年9月5日(2003.9.5)

(51)Int.Cl.⁷

G 0 6 F 17/30

識別記号

2 1 0

1 7 0

F I

G 0 6 F 17/30

テーマコード* (参考)

2 1 0 D 5 B 0 7 5

1 7 0 A

審査請求 有 請求項の数 8 O L (全 13 頁)

(21)出願番号 特願2002-46233(P2002-46233)

(22)出願日 平成14年2月22日(2002.2.22)

(71)出願人 591030237

日本ユニシス株式会社

東京都港区赤坂2丁目17番51号

(72)発明者 松田 芳雄

東京都港区赤坂二丁目17番51号 日本ユニシス株式会社内

(72)発明者 林田 英雄

東京都港区赤坂二丁目17番51号 日本ユニシス株式会社内

(74)代理人 100076428

弁理士 大塚 謙徳 (外3名)

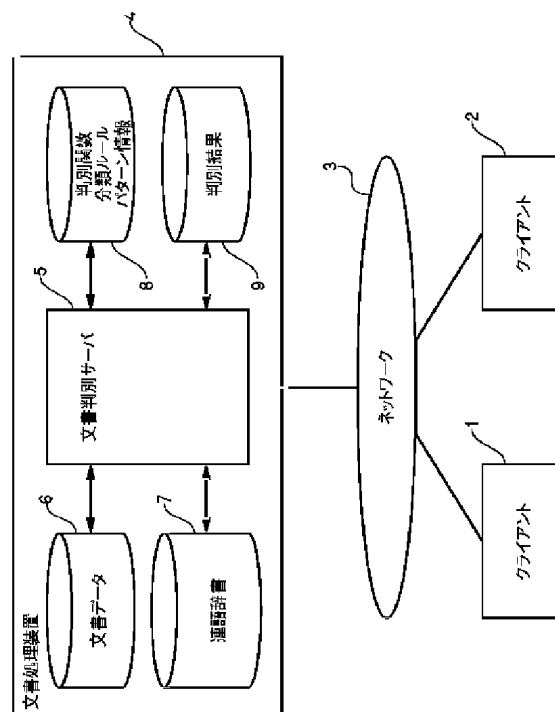
Fターム(参考) 5B075 ND03 NK32 NR12 PQ02 PQ13
PQ46

(54)【発明の名称】 情報処理装置およびその方法

(57)【要約】

【課題】 文書を機械的に判別し、結果だけを示す場合、判別精度や判別結果の信頼度が不明なため、個人に依存する判断によって、カテゴリ分けされた顧客の声が対応部門の間で盤回しされ、長期間放置される危惧がある。

【解決手段】 文書10から出現頻度が高い連語を抽出し(S20)、抽出された連語から文書のカテゴリを判別するための判別関数8aを作成し(S30)、作成された判別関数8aに基づき文書10のカテゴリを判別し(S40)、別結果の分類ルールを作成して、判別結果を分類し(S50-S60)、分類結果ごとに判別結果の正答率を計算し(S70)、判別関数、分類ルールおよび正答率を文書判別情報としてメモリに格納する。そして、文書判別情報に基づき、文書10と同類の文書6のカテゴリを判別して、その正答率を含む判別結果9を出力する(S110-S140)。



【特許請求の範囲】

【請求項1】 文書をその内容により分類する情報処理方法であって、
 所定数の文書から重要度および出現頻度に基づき連語を抽出し、
 抽出された連語から文書のカテゴリを判別するための判別関数を作成し、
 作成された判別関数に基づき、前記所定数の文書のカテゴリを判別し、
 その判別結果に基づき、前記判別結果の分類ルールを作成して、前記判別結果を分類し、
 その分類結果ごとに前記判別結果の正答率を計算し、
 前記連語、前記判別関数、前記分類ルールおよび前記正答率を文書判別情報としてメモリに格納し、
 前記文書判別情報に基づき、前記所定数の文書と同類の文書のカテゴリを判別して、その正答率を含む判別結果を出力することを特徴とする情報処理方法。

【請求項2】 前記分類ルールは、前記判別結果の平均値および標準偏差、前記判別結果を主成分分析するための固有値、並びに、前記判別結果をクラスタ分析した結果の重心であることを特徴とする請求項1に記載された情報処理方法。

【請求項3】 前記正答率は、前記判別結果および予め与えられている判別結果を示す教師信号に基づき計算されることを特徴とする請求項1または請求項2に記載された情報処理方法。

【請求項4】 前記正答率は、前記判別結果と前記教師信号とが一致する文書数の割合を示すことを特徴とする請求項3に記載された情報処理方法。

【請求項5】 前記判別結果の出力は、判別結果を各カテゴリへの帰属度の確率として表現することを特徴とする請求項1から請求項4の何れかに記載された情報処理方法。

【請求項6】 情報処理装置を制御して、請求項1から請求項5の何れかに記載された情報処理を実行することを特徴とするプログラム。

【請求項7】 請求項6に記載されたプログラムが記録されたことを特徴とする記録媒体。

【請求項8】 文書をその内容により分類する情報処理装置であって、
 所定数の文書から重要度および出現頻度に基づき連語を抽出する抽出手段と、
 抽出された連語から文書のカテゴリを判別するための判別関数を作成する作成手段と、
 作成された判別関数に基づき、前記所定数の文書のカテゴリを判別する第一の判別手段と、
 その判別結果に基づき、前記判別結果の分類ルールを作成して、前記判別結果を分類する分類手段と、
 その分類結果ごとに前記判別結果の正答率を計算する計算手段と、

前記連語、前記判別関数、前記分類ルールおよび前記正答率を文書判別情報として記憶するメモリと、
 前記文書判別情報に基づき、前記所定数の文書と同類の文書のカテゴリを判別して、その正答率を含む判別結果を出力する第二の判別手段とを有することを特徴とする情報処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は情報処理装置およびその方法に関し、例えば、文書をその内容によりカテゴリに分ける情報処理に関する。

【0002】

【従来の技術】インターネットやコールセンタなどの新しいビジネス形態の普及にともない、顧客の問い合わせ、要望および苦情などが、紙ではなく、デジタル化された文書として電子的に蓄積されることが多い。そして、各企業は、蓄積文書に記録された顧客の声を、商品企画やマーケティングへ反映させようとする。その際、顧客の声をカテゴリに分けて、各カテゴリに対応する部門に送る必要がある。

【0003】大量の文書をカテゴリに分ける処理は文書判別と呼ばれる手法で行われる。例えば、文書が予め決められた「要望」「質問」「苦情」「感想」のどのカテゴリに属するかを判別する場合は四つの判別関数を作成する。すなわち、文書が「要望」か否かを判別する関数、「質問」か否かを判別する関数、「苦情」か否かを判別する関数、および、「感想」か否かを判別する関数である。これら四つの関数によって判別確率 $P(\text{demand})$ 、 $P(\text{query})$ 、 $P(\text{complaint})$ 、 $P(\text{impression})$ が計算され、これら確率の大きさによって、文書がどのカテゴリに属するかが判別される。

【0004】

【発明が解決しようとする課題】人間が書いた文書は単純ではなく、色々な要素を含む。例えば、苦情の要素を含む「要望」とか、苦情に近い「要望」もあり得る。これらを単に「要望」「質問」「苦情」「感想」のどれかに分けることには無理があるし、判別結果の信頼度も明確ではない。

【0005】また、機械的に判別し、結果だけを示す場合、判別精度や判別結果の信頼度が不明なため、個人に依存する判断によって、カテゴリ分けされた顧客の声に対応部門の間で盤回しされ、長期間放置される危険がある。顧客の声を長期間放置するなどすれば、顧客に不安感や不信感を抱かせ、時には反感を買う危険性がある。

【0006】本発明は、上述の問題を個々にまたはまとめて解決するためのもので、高い信頼度で文書を判別することを目的とする。

【0007】また、文書の判別結果に、判別の信頼度を示す情報を与えることを他の目的とする。

【0008】

【課題を解決するための手段】本発明は、前記の目的を達成する一手段として、以下の構成を備える。

【0009】本発明にかかる情報処理方法は、文書その内容により分類する情報処理方法であって、所定数の文書から重要度および出現頻度に基づき連語を抽出し、抽出された連語から文書のカテゴリを判別するための判別関数を作成し、作成された判別関数に基づき、前記所定数の文書のカテゴリを判別し、その判別結果に基づき、前記判別結果の分類ルールを作成して、前記判別結果を分類し、その分類結果ごとに前記判別結果の正答率を計算し、前記連語、前記判別関数、前記分類ルールおよび前記正答率を文書判別情報としてメモリに格納し、前記文書判別情報に基づき、前記所定数の文書と同類の文書のカテゴリを判別して、その正答率を含む判別結果を出力することを特徴とする。

【0010】本発明にかかる情報処理装置は、文書その内容により分類する情報処理装置であって、所定数の文書から重要度および出現頻度に基づき連語を抽出する抽出手段と、抽出された連語から文書のカテゴリを判別するための判別関数を作成する作成手段と、作成された判別関数に基づき、前記所定数の文書のカテゴリを判別する第一の判別手段と、その判別結果に基づき、前記判別結果の分類ルールを作成して、前記判別結果を分類する分類手段と、その分類結果ごとに前記判別結果の正答率を計算する計算手段と、前記連語、前記判別関数、前記分類ルールおよび前記正答率を文書判別情報として記憶するメモリと、前記文書判別情報に基づき、前記所定数の文書と同類の文書のカテゴリを判別して、その正答率を含む判別結果を出力する第二の判別手段とを有することを特徴とする。

【0011】

【発明の実施の形態】以下、本発明にかかる一実施形態の文書処理を図面を参照して詳細に説明する。

【0012】〔概要〕本実施形態は、以下の方法により、顧客の声のような文書を短時間に、高い信頼度で判別するものである。

【0013】● 判別関数の精度向上

回帰分析により判別関数を作成するに当たって、使用する連語辞書を予め人が作成する。連語辞書に登録する連語とは、それぞれの判別項目に属する文書で出現頻度が高いと思われる表現である。通常、連語辞書を構成する方法として、単語形式を採用する方法、構文形式を採用する方法がある。単語形式の連語辞書は作成が容易、構文形式の連語辞書は高精度の判別関数が得られる、という特徴がある。

【0014】顧客の声のような文章を判別対象にする場合、必ずしも、正しい構文の文章とは限らない。従って、構文形式の連語辞書を採用する精度向上のメリットは得られない。そこで、本実施形態では、単語形式の連語辞書を採用し、連語辞書の内容を正規表現が可能にす

ることで、精度の高い判別関数を得る。さらに、判別関数を一次直線、二次曲線、指数曲線などから選択することで、判別関数を固定した場合に比べて、より精度の高い判別関数を得る。

【0015】● 判別結果の帰属度と正答率の提示

本実施形態では、図1に一例を示すように、判別結果を各カテゴリへの帰属度の確率として表現する。これにより、各文書が「要望」「質問」「感想」「苦情」の各要素をどの程度含むかを示し、択一ではない適切な判別を可能にする。例えば、図1の例では、感想の要素を含む「要望」、または、感想に近い「要望」と判断することができる。

【0016】そして、帰属度の確率表現と同時に、帰属度についてどの程度の信頼度があるかを正答率として提示する。

【0017】● 自動処理

本実施形態では、事前処理により、一旦、連語辞書、判別関数、分類ルールおよびパターン情報を作成し保存する。その後、同種の文書について判別を実施する際に、保存された情報を利用して、人手を介さずに自動的に判別を実施する。

【0018】本実施形態は、顧客の声のような文書を商品企画やマーケティングへ適用するために、上記の構成により、前処理として、それら文書を判別する。判別処理において、正規表現が可能な連語辞書を利用し、一次直線、二次曲線、指数曲線などから適切な判別関数を自動選択して判別精度を向上させる。そして、判別結果の正答率を示すことで、判別された文書を対応部門へ送付する際に、正答率が高い文書については自動処理し、信頼度が低い文書は人の判断を介在させるなどして、後処理（対応部門における処理）との効率的な連携が可能になる。勿論、同種の文書に対しては、保存情報を利用する自動処理によって短時間に文書判別が可能になる。

【0019】〔構成〕図2は実施形態の文書処理を実行するシステム構成例を示すブロック図である。なお、実施形態は、クライアントサーバ構成に限らず、スタンドアロンのコンピュータでも実現可能である。

【0020】図2において、文書処理の利用者は、クライアント1や2のモニタに表示された画面を介して、文書処理装置4を構成する文書判別サーバ5にアクセスする。クライアント1や2と文書判別サーバ5とはインターネットやLANなどのネットワーク3を介して接続されている。

【0021】インターネットなどを利用して収集される大量の文書データ6は、文書判別サーバ5に接続されたハードディスクなどの外部記憶装置に保存されている。利用者は、連語辞書7などを予め準備して、文書判別サーバ5に接続されたハードディスクなどの外部記憶装置に保存しておく。

【0022】〔文書判別処理〕図3は実施形態の文書判別処理の一例を示すフローチャートで、文書判別サーバ

5によって実行される処理を示す。

【0023】文書処理の利用者は、予め人手によって判別された既存の文書データ（五千件程度）を判別関数の作成用文書10として準備する。なお、判別関数の作成用文書10および文書データ6は、処理精度を上げるために、表現・表記が統一されていることが望ましい。すなわち、英字の大文字（または小文字）への統一、英字の全角文字（または半角文字）への統一、半角カナ文字を全角カナ文字への変換、並びに、意味が同じで表記が異なる文字列の表記の統一など、所謂データクリーニングを施す。

【0024】以降、図3に示す処理の流れに沿って、文書を「要望」「質問」「感想」「苦情」の四種類に判別する例を説明する。勿論、判別カテゴリはこの四つに限定されるものではなく、判別対象の文書の特性に応じたカテゴリを設定することになる。

【0025】図3に示すステップS10からS70の処理は、計算量が多く、また人手を介したり、試行錯誤による繰り返しが必要になることもあり、一般に処理に時間を要する。本実施形態は、一旦、ステップS10からS70の処理を行い、連語辞書7、判別関数8a、分類ルール8b、8cおよび8d、並びに、パターン情報8eを作成して外部記憶装置に保存する。その後、判別関数の作成用文書10に類似する内容の別の文書データ6の判別を行う場合は、図2に示すステップS110からS140の処理を自動的に実行する。従って、大量の文書データ6の判別を短時間に実行することができる。勿論、異なる分野の文書データ6を判別する場合や、時間や状況が変化して判別対象の文書の傾向に変化があると思われる場合は、改めてステップS10からS70の処理を実施すればよい。

【0026】● 連語辞書の作成(S10)

文書判別サーバ5は、利用者が予め判別した情報（以降「教師信号」と呼ぶ）に基づき連語辞書7を作成する。本実施形態においては、上述したように、単語形式の連語辞書とし、かつ、正規表現を可能にすることで、高い精度の判別関数が得られる連語辞書7にする。

【0027】図4は連語辞書を作成するための処理画面の一例を示す図で、文書判別サーバ5によってクライアント1または2のモニタに表示される。

【0028】文書判別サーバ5は、判別関数の作成用文書10の中から、あるカテゴリ、例えば要望に判別された文書を、画面左上の文書一覧21に表示する。利用者が画面右下の採用中連語一覧24に、判別のキーとなるであろう連語を入力すると、文書判別サーバ5は、その連語に一致する文字列をもつ文書の表示を、画面左上の文書一覧21から画面左下の文書一覧22へ移動する。

【0029】利用者は、複数の連語を入力・選択・取り消し（まとめて「指定」と呼ぶ）が可能である。具体的には、文書のカテゴリに一致する連語を指定する。つまり、連語は、その重要度および出現頻度などの情報に基

づき、判別関数の作成用文書10から抽出されることになる。

【0030】文書判別サーバ5は、他の連語が指定されると、上記と同様の処理を行う。また、利用者が画面右下の採用中連語一覧24に表示された連語の取消を指示すると、文書判別サーバ5は、取消対象の連語の表示を、画面右下の採用中連語一覧24から画面右上の未採用連語一覧23へ移動するとともに、取消された連語とのみ一致する文字列をもつ文書の表示を、画面左下の文書一覧22から画面左上の文書一覧21へ移動する。

【0031】本実施形態では、正規表現による次の連語指定が可能である。なお、正規表現は下記に限定されず、周知の様々な正規表現を利用することが可能である。

【0032】「*」は任意の文字列を表す。例えば「AB*CD」は文字列「AB」の後ろに一つ以上の文字があり、その後に文字列「CD」が続くことを表す。

【0033】「?」は任意の一文字を表す。例えば「AB?CD」は文字列「AB」の後ろに一文字があり、その後に文字列「CD」が続くことを表す。

【0034】「#」は数字一文字を表す。例えば「#秒以内」であれば、数字一文字の後ろに文字列「秒以内」が続くことを表す。

【0035】ブラケット（[...]）は、ブラケットに囲まれ、かつ、「,」で区切られた複数の文字列の一つを表す。例えば「[証明,立証]する」は「証明する」または「立証する」を表す。

【0036】利用者が指定した連語から、文書判別サーバ5は連語辞書7を作成し、外部記憶装置に保存する。図5は連語辞書7の例を示す図である。

【0037】図5において「判別項目番号」は判別カテゴリを整数で示す。例えば、判別項目番号「1」は「要望」を、「2」は「質問」を、「3」は「感想」を、「4」は「苦情」をそれぞれ示す。「採用フラグ」は、その連語が採用されているか否かを示し「1」は採用、「0」は不採用を示す。

【0038】なお、文書判別サーバ5は、ステップS10の処理を再実行する場合、前回作成した連語辞書7を読み込んで図4に示す処理画面をクライアントのモニタに表示する。従って、利用者は処理画面を参照して連語辞書7の修正を指示することが可能である。

【0039】● 連語の抽出(S20)

文書判別サーバ5は、判別関数の作成用文書10の一文書ごとに、連語辞書7に登録された全連語について、該当する連語を含むか否かを調べ、図6に示すテーブルを作成する。

【0040】図6に示すテーブルにおいて、「教師信号」の列は各行の文書の該当カテゴリを示し、教師信号「2」は対応する文書が該当カテゴリに分類されたことを、「1」は対応する文書が該当カテゴリに分類されて

いないことを示す。例えば、文書No.1は「要望」カテゴリに分類されている。「連語の出現」の列は、各行の文書に該当する連語が出現するか否かを示し「1」は該当する連語が出現することを、「0」は該当する連語が出現しないことを示す。

【0041】● 判別関数の作成(S30)

$$y = b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_n \cdot x_n + c \quad \dots(1)$$

ここで、 y は「教師信号」の値(図6参照)

x は「連語の出現」の値(図6参照)

n はその判別における連語の数

【0042】文書判別サーバ5は、図6に示す連語の抽出結果を示すテーブルの値に対して回帰分析を行い、定数 b_1 、 b_2 、 b_3 、 \dots 、 b_n および c の各値を計算する。回帰分析については、例えば、田中豊、脇本和昌「多変量統計解析法」(現代数学社、1983年5月、第1章回帰分析法)に示されている方法などを利用する。

【0043】文書判別サーバ5は、回帰式を作成した後、曲線を当て嵌めて式を補正する。文書判別サーバ5が行う曲線の当て嵌めにあたり、カテゴリ「要望」「質問」「感想」「苦情」それぞれについて、図7に示す処理を行う。

【0044】連語の抽出結果を示すテーブルの全データを、「連語の出現」の値に基づき、回帰式に当て嵌めて、教師信号に相当する値の予測値(以降「スコア」と呼ぶ)を計算する(S301)。

【0045】スコアの値をキーにして文書をソートする(S302)。

【0046】ソート結果の文書を自然数 m (例えば $m=20$) のグループに分ける(S303)。ただし、同一スコアの値をもつ文書は同一グループ(以降「ランク」と呼ぶ)に分けるため、各ランクの文書の数が同一になるとは限らない。

【0047】各ランクのスコアの値の平均値(以降「スコア平均」と呼ぶ)を計算する。また、ランクごとに、連語抽出結果を示すテーブル(図6)の「教師信号」の値が「2」である文書の数(以降「反応数」と呼ぶ)の割合(以降「反応率」と呼ぶ)を算出する(S304)。そし

$$\log Y = (A + BX + CX^2) \cdot \log 10$$

$$= \log 10 \cdot A + \log 10 \cdot BX + 2 \cdot \log 10 \cdot CX \quad \dots(3)$$

【0054】続いて、 m 個のスコア平均を使用して、反応率が予測されるように曲線に当て嵌める(S307)。当て嵌める曲線は例えば次の四種類である。

一次直線: $Y = A + BX$

二次曲線: $Y = A + BX + CX^2$

指数曲線: $Y = AB^X$

二次指数曲線: $Y = 10^Z$

ここで、 $Z = A + BX + CX^2$

【0055】上記の曲線の何れにおいても、 Y は反応率であり、 X は各ランクのスコア平均である。文書判別サーバ5は、ステップS306と同様の方法で回帰分析を行

文書判別サーバ5は、四つのカテゴリに対応する判別関数を回帰分析の手法で作成する。すなわち「要望」か否かを判別する関数、「質問」か否かを判別する関数、「感想」か否かを判別する関数、および、「苦情」か否かを判別する関数を作成する。具体的には回帰式(1)が作成される。

て、各ランクの「教師信号」の値の平均値(以降「評価変数平均」と呼ぶ)を計算する(S305)。

【0048】図8は以上の処理結果の一例を示す図である。なお「サンプル数」列は各ランクに含まれる文書の数を示す。

【0049】次に、 m 個のスコア平均を使用して、評価変数が予測されるように曲線を当て嵌める(S306)。当て嵌める曲線は例えば次の四種類である。

一次直線: $Y = A + BX$

二次曲線: $Y = A + BX + CX^2$

指数曲線: $Y = AB^X$

二次指数曲線: $Y = 10^Z$

ここで、 $Z = A + BX + CX^2$

【0050】上記の曲線の何れにおいても、 Y は評価変数であり、 X は各ランクのスコア平均である。文書判別サーバ5は、以下に示す方法で回帰分析を行い、 A 、 B および C の値を計算する。

【0051】一次曲線および二次曲線の場合、スコア平均の値を X に代入して回帰分析を行い、 A 、 B および C を求める。

【0052】指数曲線の場合は、その両辺を対数変換した式(2)に評価変数平均の値から求まる $\log Y$ 、および、スコア平均の値 X を代入して回帰分析を行い、 A および B を求める。

$$\log Y = \log A + X \cdot \log B \quad \dots(2)$$

【0053】二次指数曲線の場合は、両辺を対数変換した式(3)に評価変数平均の値から求まる $\log Y$ 、並びに、スコア平均の値 X から求まる $\log X$ および $2 \cdot \log X$ を代入して回帰分析を行い、 A 、 B および C を求める。

い、 A 、 B および C の値を計算する。

【0056】次に、文書判別サーバ5は、ステップS306およびS307でそれぞれ得られた四種類の曲線をクライアントの画面に並べてグラフ表示する(S308)。図9はステップS306で得られる評価変数の予測結果を示す図、図10はステップS307で得られる反応率の予測結果を示す図である。なお、図11に示す数値表示も可能である。さらに、文書判別サーバ5は、採用を推奨する曲線の一つ選定して、図9および10に示すように推奨曲線のグラフを囲んだり、図11に示すように推奨曲線の採用フラグを「1」にして、推奨曲線を示すことができる。

【0057】文書判別サーバ5は、採用を推奨する曲線の選定に当り、決定係数Rの高いものを選定する。決定係数とは、回帰分析による式の当て嵌りの度合を表す指標で、0から1の間の値をとる。なお、利用者はグラフ表示または数値表示を観て、採用する曲線の変更を指示することもできる。また、採用される曲線の種類はカテゴリ「要望」「質問」「苦情」「感想」ごとに異なってもよく、例えば「要望」は二次曲線を、「質問」は指数曲線を採用することができる。

【0058】文書判別サーバ5は、採用された曲線の数値情報を各カテゴリの判別関数8aとして外部記憶装置に保存する(S309)。

【0059】● 判別関数を使用した文書判別(S40)
文書判別サーバ5は、図6に示した連語抽出結果を判別関数8aに当て嵌めて、判別関数の作成用文書10の各文書の判別関数の値を計算する(以降、計算結果を「判別結果」と呼ぶ)。なお、以下では、反応率を予測する判別関数を例にして説明する。

【0060】図12は判別結果の一例を示す図である。図12において「教師信号」列は各行の文書の該当カテゴリを示し、教師信号「1」は対応する文書が該当カテゴリに分類されたことを、「0」は対応する文書が該当カテゴリに分類されていないことを示す。また、「反応率」列は判別関数の値の計算結果を示す。そして、「判別結果」列は、該当する文書の該当するカテゴリの反応率が閾値(例えば50)以上の場合に「1」になる。

【0061】● 主成分分析によるマップ作成(S50)
文書判別サーバ5は、前のステップS40で計算した文書の判別結果を基準化し、多変量解析の手法の一つである主成分分析を用いて、判別結果のデータを空間上に配置する。基準化は、図13に示す各データの平均値(例えば「要望」の判別結果の平均値)とその標準偏差(例えば「要望」の判別結果の標準偏差)を用いて、田中豊、脇本和昌「多変量統計解析法」(現代数学社、1983年5月、第2章主成分分析法)に示されるような方法(式(4)参照)によって行う。

$$Dstd = (D - Dave) / Dsd \quad \dots(4)$$

ここで、Dstd: 基準化後の判別結果

D: 基準化前の判別結果

Dave: 判別結果の平均値

Dsd: 判別結果の標準偏差

【0062】そして、主成分分析の結果できあがる空間上に、各判別結果を配置してサンプルマップを作成する。サンプルマップの作成も、前記の文献に示されるような方法を利用する。

【0063】図14および図15は作成されるサンプルマッ

同一分類内で教師信号と判別結果が一致した文書数

$$\text{正答率} = \frac{\text{同一分類内で教師信号と判別結果が一致した文書数}}{\text{同一分類内の文書数}} \times 100 \quad \dots(b)$$

同一分類内の文書数

【0069】文書判別サーバ5は、ここまでの処理で外

部のグラフ表示例および数値表示例を示す図である。また、図16は、基準化された判別結果を主成分分析の結果できあがる空間上に配置するために、主成分分析で利用される固有値の例を示す図である。さらに、文書判別サーバ5は、判別結果の基準化に用いた、図13に示される、各判別結果の平均と標準偏差を分類ルール8bとして、図16に示される固有値を分類ルール8cとして外部記憶装置に保存する。

【0064】● クラスタ分析による判別結果の分類(S60)

文書判別サーバ5は、前のステップS50で作成したサンプルマップについて、多変量解析の手法の一つであるk-means法によるクラスタ分析を行い、判別結果のデータをグループ化する。なお、この処理を「分類」と呼び、グループの総称を「クラスタ」と呼ぶ。

【0065】文書判別サーバ5は、得られたクラスタに属する全サンプルの重心を計算し、その結果のクラスタの重心(図17参照)を分類ルール8dとして外部記憶装置に保存する。なお、クラスタ分析については、例えば、西田英郎監訳「クラスタ分析とその応用」(株式会社内田老鶴園、1995年12月、7章 非階層的クラスタ分析法)に紹介されている。

【0066】サンプルマップのすべての軸についてk-means法によるクラスタ分析を行うと、すべての軸を用いた計算を必要とし、計算に長時間を要する。このため、クラスタ分析に使用される軸の数は、出力されるすべての軸の数ではなく、各軸のサンプルのばらつきなどを観ながら、試行錯誤を繰り返して軸数mを決定する。このため、データを分類するたびにK-means法によるクラスタ分析を行えば、非常に手間がかかる処理になる。そこで、本実施形態では、これを改善するために、後述するように、二回目以降の分類実施処理においてはk-means法によるクラスタ分析を行わずに分類結果を生成する。

【0067】● 判別結果の正答率の計算(S70)

文書判別サーバ5が計算した反応率は、図12に示すように、0から100の間の実数値をとるので、判別結果の値について閾値(例えば50)で真偽を求め、事前に利用者が与えた教師信号と突き合わせて、正答か否かを判定する。そして分類ごとの正答率(図18参照)をパターン情報8eとして外部記憶装置に保存する。

【0068】ここで、正答率は式(5)で示される。例えば、図18において、分類番号1の文書数が579件、そのうち教師信号と判別結果が一致した文書数は553件であるから、分類番号1の正答率は $553/579 \times 100 = 95.51\%$ になる。

部記憶装置に保存された連語辞書7、判別関数8a、分類

ルール8b（各判別結果の平均と標準偏差）、分類ルール8c（固有値）、分類ルール8d（クラスタの重心）、および、パターン情報8eを利用することで、この後、同じような内容の文書があれば、何時でも人手を介すること無く、図3に示すステップS110からS140の手順を自動的に実行し、文書判別を行うことができる。例えば、文書判別サーバ5は、毎日集まってくるデータを、毎日、自動的に判別することができる。

【0070】続いて、図3に示すステップS110からS140の処理を説明する。

【0071】● 連語の抽出(S110)

文書判別サーバ5は、判別対象の文書データ6の文書一件ごとに、連語辞書7の全連語について、該当する連語を含むか否かを調べて、図19に示すテーブルを作成する。

【0072】● 判別関数を使用する判別(S120)

文書判別サーバ5は、図19に示す連語抽出結果を示すテーブルについて、文書データ6の文書一件ごとに「要望」「質問」「苦情」「感想」の四つのカテゴリに対応する判別関数を用いて、判別関数の値（判別結果）を計算する。

【0073】図20は判別結果のテーブル例を示す図で、「確率」列は該当する文書の各判別関数の計算結果の値を示す。「判別結果」列は該当する文書の、該当するカテゴリの「確率」の値がある閾値（例えば50）以上であれば「1」になり、閾値未満であれば「0」になる。

【0074】●判別結果のマップ上への配置と分類(S130)

文書判別サーバ5は、分類ルール8b（平均および標準偏差）、分類ルール8c（固有値）、分類ルール8d（クラスタの重心）を読み込み、次の手順で、判別結果をマップ上へ配置し分類を作成する。

(1) 図20に示される判別結果テーブルから文書データ、例えば文書No. 30を取り出し、その確率に基づき、分類ルール8bを使用して、ステップS50と同様の基準化を行う（式(4)参照）。図21は基準化後の確率の一例を示す図である。

(2) 基準化された確率に基づき、分類ルール8cを使用して、以下のような方法で、マップ上の位置Zを計算し、文書データ6をマップ上に配置する。図22はマップ上の位置Zの例を示す図である。

$$Z = d1 \cdot d2 + q1 \cdot q2 + i1 \cdot i2 + c1 \cdot c2$$

ここで、d1: 基準化された要望予測値[率]

d2: マップ（数値情報）の軸1の要望予測値[率]

q1: 基準化された質問予測値[率]

q2: マップ（数値情報）の軸1の質問予測値[率]

i1: 基準化された感想予測値[率]

i2: マップ（数値情報）の軸1の感想予測値[率]

c1: 基準化された苦情予測値[率]

c2: マップ（数値情報）の軸1の苦情予測値[率]

(3) 図17に示すクラスタの重心の各分類の重心と、位置

Zとの間の距離を計算する。

(4) 距離が最小となる分類の番号を「最も近い分類番号」として記憶する。

(5) 以降、すべての文書データについて、上記(1)から(4)の処理を繰り返す。

【0075】このように文書判別サーバ5は、分類ルール8bから8dを利用することで、文書データ6に対して自動的に分類処理を行うことができる。つまり、異なる文書データに対して、毎回、判別関数の作成、主成分分析およびクラスタ分析を人手を介して行う必要がない。

【0076】図23は文書判別サーバ5が図20に示す判別結果の分類を作成した結果の一例を示す図である。

【0077】また、マップの作成(S50)における主成分分析に要する処理時間、および、K-means法によるクラスタ分析(S60)に要する処理時間を削減できるため、短時間に分類結果を得ることができる。

【0078】● 判別結果9の出力(S140)

分類の作成結果から得られた「最も近い分類番号」に基づき、図18に示すパターン情報から得られる「正答率」を、判別結果9に付加して出力する。

【0079】図24は正答率付きの判別結果9の一例を示す図である。

【0080】このように、本実施形態によれば、文書判別サーバ5は、最初の処理（S10からS70）で、連語辞書7、判別関数8a、分類ルール8b（各平均と標準偏差）、分類ルール8c（マップの数値情報）、分類ルール8d（クラスタの重心）およびパターン情報8eを作成して外部記憶装置に保存する。つまり、内容が類似する文書の場合、経験者による人手処理を介させたステップS10からS70の繰り返し処理が不要になる。例えばインターネットを利用したアンケート収集のように、毎日、同じような内容の文書データ6が集まる場合、一度、ステップS10からS70の処理を行えば、以降、ステップS110からS140の処理により、逐次発生する文書データ6を自動的に文書判別処理することができる。

【0081】さらに、文書データ6の判別結果9に正答率を付加するため、その判別精度や判別結果の信頼度が明確になる。例えば、正答率の高い、カテゴリ分けされた顧客の声は適切な部門へ自動送付して対応を図ることができるし、逆に、正答率の低い、カテゴリ分けされた顧客の声は、その詳細を分析するように担当者へ送付するなどができる。従って、顧客の声が対応部門の間で盤回しされたり、長期間放置されるなどを防ぐことができる。

【0082】

【発明の効果】以上説明したように、本発明によれば、高い信頼度で文書を判別することができる。

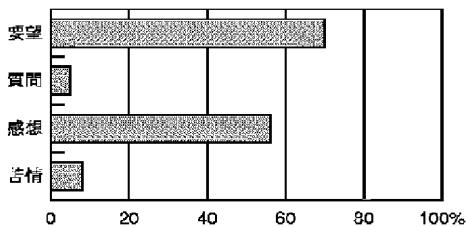
【0083】また、文書の判別結果に、判別の信頼度を示す情報を与えることができる。

【図面の簡単な説明】

- 【図1】カテゴリの判別結果の表現例を示す図、
 【図2】文書処理を実行するシステム構成例を示すブロック図、
 【図3】文書判別処理の一例を示すフローチャート、
 【図4】連語辞書を作成するための処理画面の一例を示す図、
 【図5】連語辞書の例を示す図、
 【図6】連語の抽出を説明する図、
 【図7】判別関数の作成を説明するフローチャート、
 【図8】判別関数の作成を説明する図、
 【図9】評価変数の予測結果を示す図、
 【図10】反応率の予測結果を示す図、
 【図11】評価変数と反応率の予測結果を数値表現で示す図、
 【図12】文書の判別結果の一例を示す図、
 【図13】判別結果の平均値および標準偏差の一例を示す図、

- す図、
 【図14】サンプルマップのグラフ表示例を示す図、
 【図15】サンプルマップの数値表示例を示す図、
 【図16】主成分分析で使用する固有値の例を示す図、
 【図17】クラスタの重心を説明する図、
 【図18】分類ごとの正答率を説明する図、
 【図19】連語の抽出結果を示すテーブル例を示す図、
 【図20】判別結果のテーブル例を示す図、
 【図21】基準化後の確率の一例を示す図、
 【図22】判別結果のマップ上への配置を説明する図、
 【図23】判別結果の分類を作成した結果の一例を示す図、
 【図24】正答率付きの判別結果の一例を示す図である。

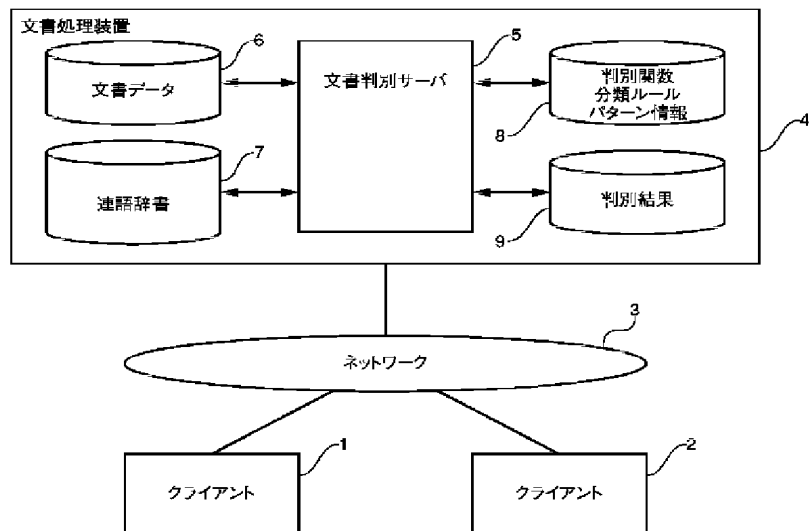
【図1】



【図5】

| 連番 | 判別項目番号 | 採用フラグ | 連語 |
|-------|--------|-------|------------|
| 3 | 1 | 1 | [前,昔] |
| 4 | 1 | 1 | あったら |
| 5 | 1 | 1 | あったらうれしい |
| 6 | 1 | 1 | あっても |
| 7 | 1 | 1 | ありますか? |
| 8 | 1 | 1 | あると |
| 9 | 1 | 1 | いかがですか |
| 10 | 1 | 1 | いたい |
| 11 | 1 | 1 | いただ |
| 12 | 1 | 0 | うれし*[おも,思] |
| 13 | 1 | 1 | お願い |
| 14 | 1 | 1 | きたい |
| 15 | 1 | 1 | ください |
| 16 | 1 | 1 | さーい |
| 17 | 1 | 1 | したい |
| 18 | 1 | 1 | して |
| 19 | 1 | 1 | すが |
| 20 | 1 | 1 | せんか |
| 21 | 1 | 1 | ぜび |
| 22 | 1 | 1 | たいない |
| 23 | 1 | 1 | たら |
| 24 | 1 | 1 | だった |
| 25 | 1 | 1 | できませんか |
| 26 | 1 | 1 | でしょうか |
| 27 | 1 | 1 | ではないかしら |
| 28 | 1 | 1 | と、いい |
| | | | ~~ |

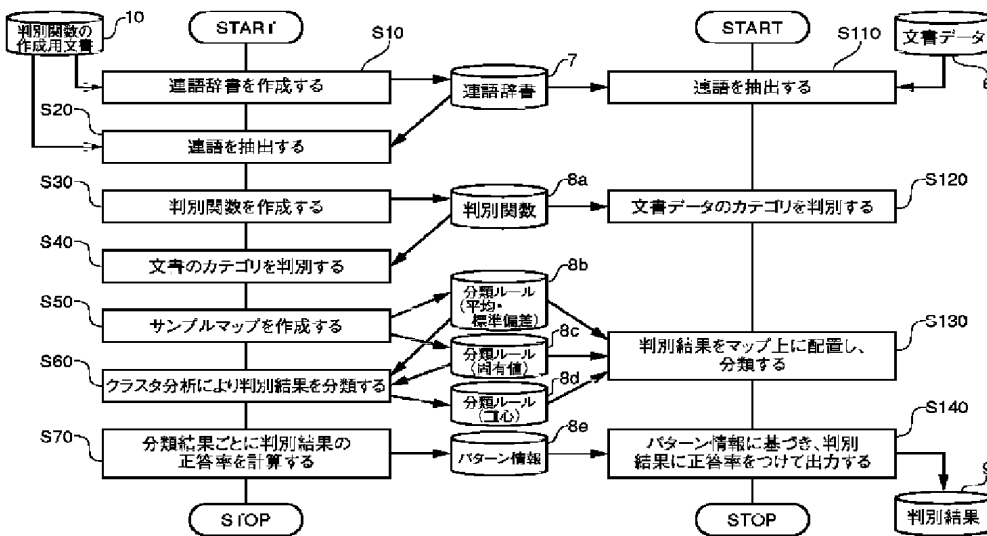
【図2】



【図11】

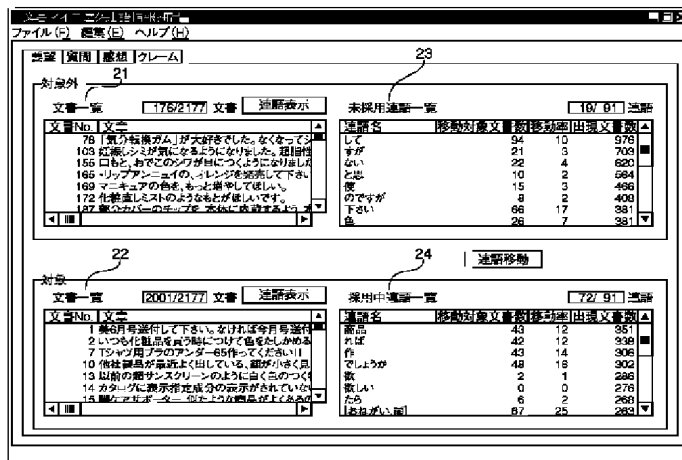
| 回帰予測値 (ベコア) | A | B1 | B2 | 決定係数 (R) | 採用 |
|-------------------------------|-----------|----------|----------|----------|----|
| $Y = A + B1X1 + B2X2 + \dots$ | 1.3492 | -0.04361 | 0.1827 | 0.67127 | |
| 評価変数予測 | A | B | C | | |
| $Y = A + BX$ | -0.0250 | 1.0174 | | 0.95 | 0 |
| $Y = A + BX + CX^2$ | -0.4280 | 1.5793 | -0.1855 | 0.95 | 0 |
| $Y = AB^X$ | -0.6749 | 0.7012 | | 0.95 | 0 |
| $Y = 10(A + BX + CX^2)$ | -0.5154 | 0.6138 | -0.1023 | 0.96 | 0 |
| 反応率予測 | A | B | C | | |
| $Y = A + BX$ | -102.5034 | 101.7414 | | 0.95 | 0 |
| $Y = A + BX + CX^2$ | -142.7990 | 157.8307 | -18.5464 | 0.95 | 0 |
| $Y = AB^X$ | 3.3570 | -1.5880 | | 0.92 | 0 |
| $Y = 10(A + BX + CX^2)$ | -3.7998 | -5.7870 | -1.4315 | 0.98 | 1 |

【図3】



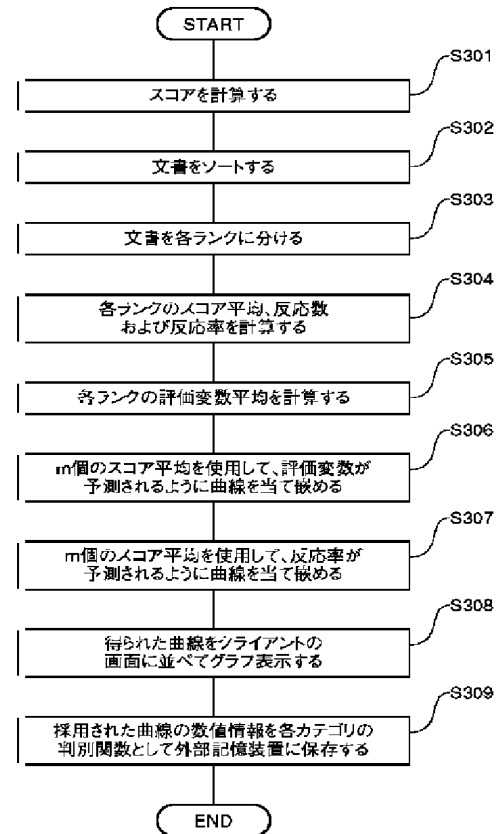
【図4】

【図7】



【図15】

| | 軸1 | 軸2 | 軸3 | 軸4 |
|----|-------|-------|-------|-------|
| 1 | 1.55 | -0.07 | -0.18 | -0.28 |
| 2 | -0.14 | -1.16 | -0.34 | 0.71 |
| 3 | -0.81 | 0.18 | -0.77 | -0.41 |
| 4 | -0.85 | 0.33 | 1.17 | -0.34 |
| 5 | -0.32 | 0.93 | -1.05 | -0.12 |
| 6 | -2.04 | -0.47 | -0.37 | -0.05 |
| 7 | 1.45 | -0.41 | -0.34 | -0.38 |
| 8 | 1.03 | 0.74 | 0.75 | 0.13 |
| 9 | -1.76 | 2.21 | -1.43 | 0.24 |
| 10 | 2.29 | -0.66 | -0.34 | -0.18 |
| 11 | 0.21 | 1.01 | 0.4 | 0 |
| 12 | -0.53 | 2.22 | -1.65 | -0.08 |
| 13 | 1.27 | 0.62 | -1.36 | 0.92 |
| 15 | 1.04 | 0.39 | -0.23 | -0.21 |
| 16 | 1.62 | -0.07 | -0.71 | 0.07 |
| 17 | 1.69 | -0.92 | -0.32 | -0.1 |



【図6】

| 文書No. | 教師信号 | | | | 連語の出現 | | | | | |
|-------|------|----|----|----|----------|-----|---------|-----|-----|-------|
| | 要望 | 質問 | 感想 | 苦情 | いと[思,おも] | あると | [やす,安]い | いない | お願い | ～～ |
| 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 3 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 4 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | |
| 5 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | |
| 6 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | |
| 7 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 8 | 1 | 1 | 1 | 2 | 0 | 1 | 0 | 1 | 0 | |
| 9 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | |
| 10 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 11 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 12 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | |
| 13 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 14 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | |
| 15 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | |
| 16 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | |

【図8】

| ランク | サンプル数 | メコア平均 | 反応数 | 反応率 | 評価変数平均 |
|-----|-------|-------|-----|-------|--------|
| 1 | 243 | 2.20 | 240 | 96.77 | 1.97 |
| 2 | 247 | 1.95 | 236 | 95.55 | 1.96 |
| 3 | 247 | 1.64 | 228 | 92.31 | 1.92 |
| 4 | 247 | 1.77 | 224 | 90.69 | 1.91 |
| 5 | 247 | 1.70 | 219 | 88.66 | 1.89 |
| 6 | 247 | 1.64 | 187 | 75.71 | 1.76 |
| 7 | 247 | 1.58 | 173 | 70.04 | 1.70 |
| 8 | 247 | 1.52 | 157 | 63.56 | 1.64 |
| 9 | 247 | 1.46 | 109 | 44.13 | 1.44 |
| 10 | 247 | 1.42 | 102 | 41.30 | 1.41 |
| 11 | 247 | 1.37 | 75 | 30.36 | 1.30 |
| 12 | 247 | 1.33 | 59 | 23.89 | 1.24 |
| 13 | 247 | 1.29 | 40 | 16.19 | 1.16 |
| 14 | 247 | 1.25 | 36 | 14.57 | 1.15 |
| 15 | 247 | 1.22 | 32 | 12.96 | 1.13 |
| 16 | 247 | 1.17 | 16 | 6.48 | 1.06 |
| 17 | 247 | 1.13 | 17 | 6.88 | 1.07 |
| 18 | 247 | 1.08 | 12 | 4.86 | 1.05 |
| 19 | 247 | 1.01 | 8 | 3.24 | 1.03 |
| 20 | 247 | 0.87 | 6 | 2.43 | 1.02 |

【図13】

| | 要望予測値[率] | 質問予測値[率] | 感想予測値[率] | 苦情予測値[率] |
|------|----------|----------|----------|----------|
| 平均 | 44.41 | 16.88 | 29.00 | 25.22 |
| 標準偏差 | 30.44 | 23.64 | 28.85 | 25.62 |

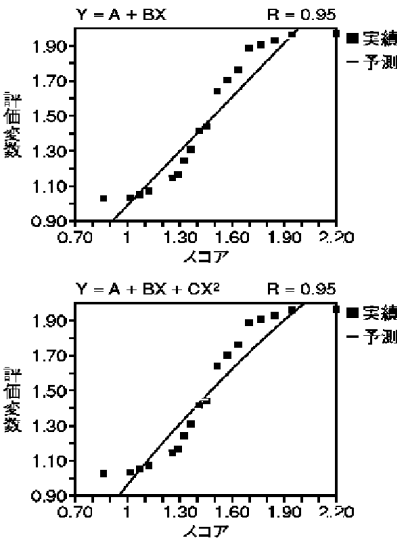
【図16】

| | 軸1 | 軸2 | 軸3 |
|----------|--------|--------|--------|
| 要望予測値[率] | 0.769 | -0.189 | -0.132 |
| 質問予測値[率] | -0.061 | 0.322 | 0.868 |
| 感想予測値[率] | -0.579 | -0.606 | -0.051 |
| 苦情予測値[率] | -0.264 | 0.703 | -0.476 |

【図12】

| 文書No. | 教師信号 | | | | 反応率 | | | | 判別結果 | | | |
|-------|------|----|----|----|--------|-------|-------|-------|------|----|----|----|
| | 要望 | 質問 | 感想 | 苦情 | 要望 | 質問 | 感想 | 苦情 | 要望 | 質問 | 感想 | 苦情 |
| 1 | 1 | 0 | 0 | 0 | 76.75 | 7.86 | 0.00 | 12.44 | 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 64.76 | 4.19 | 71.16 | 9.55 | 1 | 0 | 1 | 0 |
| 7 | 1 | 0 | 0 | 0 | 75.07 | 1.30 | 6.52 | 7.77 | 1 | 0 | 0 | 0 |
| 10 | 1 | 0 | 0 | 0 | 100.00 | 0.00 | 0.00 | 0.00 | 1 | 0 | 0 | 0 |
| 13 | 1 | 0 | 0 | 0 | 92.76 | 0.00 | 13.34 | 55.06 | 1 | 0 | 0 | 1 |
| 14 | 1 | 0 | 0 | 0 | 63.60 | 11.74 | 1.74 | 25.59 | 1 | 0 | 0 | 0 |
| 15 | 1 | 0 | 0 | 0 | 86.95 | 0.00 | 5.34 | 22.44 | 1 | 0 | 0 | 0 |
| 16 | 1 | 0 | 0 | 0 | 88.73 | 0.00 | 15.76 | 0.00 | 1 | 0 | 0 | 0 |

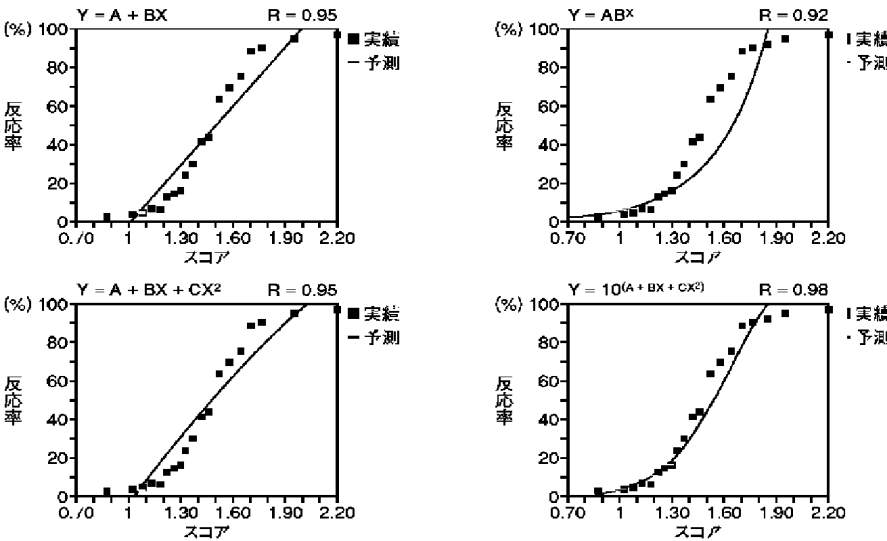
【図9】



【図21】

| 文書No. | 基準化後の確率 | | | |
|-------|---------|-------|-------|-------|
| | 要望 | 質問 | 感想 | 苦情 |
| 30 | 1.12 | -0.17 | -0.44 | -0.78 |
| 31 | 1.20 | -0.69 | -0.74 | -0.34 |
| 32 | -1.46 | 3.37 | -0.66 | 1.40 |
| 33 | 0.11 | -0.64 | -0.11 | 0.97 |
| 34 | -1.06 | -0.68 | -0.87 | 1.74 |
| 35 | 1.05 | -0.64 | -0.65 | -0.05 |
| 36 | 0.87 | -0.71 | -0.56 | 0.65 |
| 37 | -1.26 | -0.71 | -0.93 | 2.63 |
| 38 | -1.46 | -0.48 | 0.65 | 0.81 |
| 39 | -1.06 | 1.59 | 0.29 | -0.29 |

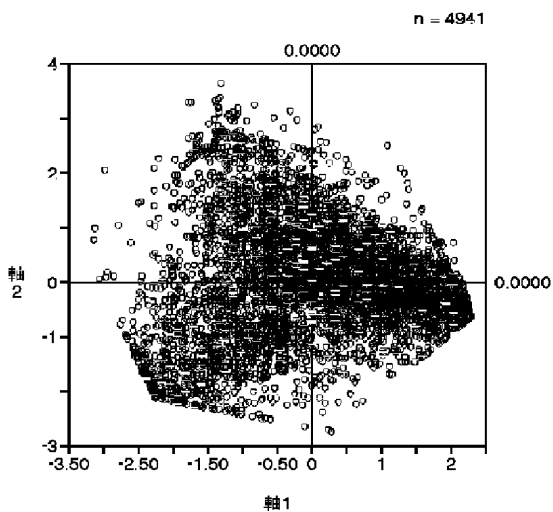
【図10】



【図17】

| | 分類1 | 分類2 | 分類3 | 分類4 | 分類5 | 分類6 | 分類7 | 分類8 | 分類9 | 分類10 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|-------|--------|
| サンプル数 | 579 | 430 | 674 | 415 | 137 | 636 | 1094 | 223 | 463 | 290 |
| 軸1 | -1.721 | -1.079 | -0.213 | -0.972 | -1.295 | 0.549 | 1.586 | -0.564 | 0.640 | -0.241 |
| 軸2 | -1.576 | 0.291 | -0.894 | 2.129 | -0.287 | 0.402 | -0.426 | 1.422 | 0.304 | 1.032 |
| 軸3 | -0.112 | -0.670 | -0.134 | -1.404 | 1.508 | -0.606 | -0.286 | 0.869 | 1.072 | 2.855 |

【図14】



【図20】

| 文書No. | 確率 | | | | 判別結果 | | | |
|-------|-------|-------|-------|-------|------|----|----|----|
| | 要望 | 質問 | 感想 | 苦情 | 要望 | 質問 | 感想 | 苦情 |
| 30 | 81.20 | 12.74 | 16.30 | 5.32 | 1 | 0 | 0 | 0 |
| 31 | 80.82 | 0.60 | 7.59 | 16.58 | 1 | 0 | 0 | 0 |
| 32 | 0.00 | 96.52 | 8.89 | 61.13 | 0 | 1 | 0 | 1 |
| 33 | 47.83 | 1.76 | 25.69 | 50.03 | 0 | 0 | 0 | 1 |
| 34 | 12.22 | 0.77 | 3.98 | 69.86 | 0 | 0 | 0 | 1 |
| 35 | 76.30 | 1.67 | 10.19 | 24.01 | 1 | 0 | 0 | 0 |
| 36 | 71.02 | 0.00 | 12.84 | 41.98 | 1 | 0 | 0 | 0 |
| 37 | 6.07 | 0.00 | 2.13 | 92.51 | 0 | 0 | 0 | 1 |
| 38 | 0.00 | 5.65 | 47.85 | 45.99 | 0 | 0 | 0 | 0 |
| 39 | 12.05 | 54.53 | 37.32 | 17.68 | 0 | 1 | 0 | 0 |

【図18】

| 分類番号 | 各分類で教師信号と一致した文書数 | 各分類の文書数 | 正答率 | 件数構成比 | 累積構成比 |
|------|------------------|---------|--------|-------|-------|
| 1 | 553 | 579 | 95.51% | 12% | 12% |
| 7 | 1035 | 1094 | 94.51% | 22% | 34% |
| 4 | 377 | 415 | 90.84% | 8% | 42% |
| 10 | 249 | 290 | 85.86% | 6% | 48% |
| 6 | 317 | 636 | 49.84% | 13% | 61% |
| 5 | 68 | 137 | 49.64% | 3% | 64% |
| 3 | 313 | 674 | 46.44% | 14% | 77% |
| 8 | 92 | 223 | 41.26% | 5% | 82% |
| 9 | 184 | 463 | 39.74% | 9% | 91% |
| 2 | 160 | 430 | 37.21% | 9% | 100% |

全文書数 4941

【図19】

| 文書No. | 文章 | いいと思おもい | あると | やす安い | いない | お嫌い |
|-------|---|---------|-----|------|-----|-----|
| 30 | コンフォートサポートソックスにクリーム色・無縫シヨーツに シャンパンはできませんか。 | 0 | 0 | 0 | 0 | 0 |
| 31 | ボディ美白ローションの詰替えを出して欲しいです。 | 0 | 0 | 0 | 0 | 0 |
| 32 | 前回顔用サンスクリーンを買って常時使用していますが、 つけたあとベタつきがあり、オイルリムーバで油分を取ってか らパウダーファンデーションをつけています。これはオイルカッ トではないのですか以後もこの方法で使用してよろしいです。 | 0 | 0 | 0 | 0 | 0 |
| 33 | (1) 顔用のソックスを愛用していましたが、口ゴムはきつすぎ と思います。20KMの夏ハイキングに使用したら、ぬいだ時 の足首の赤いあととただれがしっかりつき、数日はメカートが はけませんでした。自宅用にも不向き。(2) スポーツブラをもう 一度検討して下さい。 | 0 | 0 | 0 | 0 | 0 |
| 34 | 4121 相模コーゲン1本のみ破損(化粧箱底が割れて やぶれていた) 賞味期限00.6.3 返品交換済 | 0 | 0 | 0 | 0 | 0 |
| 35 | クレンジング剤やサンスクリーン剤、シャンプーはミニサイ ズがあると良いですね。旅行の度に何個もボトルを用意して詰 め替えるのは手間なので……。 | 0 | 0 | 0 | 0 | 0 |
| 36 | プレストパウダー化粧直し用のケースがもう少し薄くスリムに なるといいと思います他のファンデーションケースよりも厚く、 小さなバッグの時に重なります。 | 1 | 1 | 0 | 0 | 0 |
| 37 | グッドサポートストッキングPB一度の洗濯で変色 | 0 | 0 | 0 | 0 | 0 |
| 38 | 注文サンプルと間違っ入ってましたお返しします足立 | 0 | 0 | 0 | 0 | 0 |
| 39 | UV CUT57をつけて毎年ゴルフをしても日焼けせず安心して いましたが以前の長いチューブ入りの方が日焼けしなかった ように思いますか?何時も、UV CUT57+UV CUT40をつ けています。ゴルフ大好きの60才台の主婦には「美白BB」 | 0 | 0 | 0 | 0 | 0 |

【 図 2 2 】

| 文書No. | 軸上の位置 | | |
|-------|-------|-------|-------|
| | 軸1 | 軸2 | 軸3 |
| 30 | 1.40 | -0.56 | 0.08 |
| 31 | 1.48 | -0.74 | -0.58 |
| 32 | -1.32 | 2.74 | 2.48 |
| 33 | -0.06 | 0.52 | -1.03 |
| 34 | -0.73 | 1.73 | -1.24 |
| 35 | 1.23 | -0.04 | -0.64 |
| 36 | 0.87 | 0.40 | -1.02 |
| 37 | -1.08 | 2.42 | -1.66 |
| 38 | -1.69 | 0.30 | -0.64 |
| 39 | -1.00 | 0.33 | 1.65 |

【 図 2 3 】

| 文書No. | 重心への 距離 (分類1) | 重心への 距離 (分類2) | 重心への 距離 (分類3) | 重心への 距離 (分類4) | 重心への 距離 (分類5) | 重心への 距離 (分類6) | 重心への 距離 (分類7) | 重心への 距離 (分類8) | 重心への 距離 (分類9) | 重心への 距離 (分類10) | 最も近い 分類番号 |
|-------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|----------------------|--------------|
| 30 | 3.287 | 2.728 | 1.660 | 3.883 | 3.062 | 1.459 | 0.433 | 2.902 | 1.521 | 3.597 | 7 |
| 31 | 3.498 | 2.614 | 1.864 | 3.510 | 3.459 | 1.129 | 0.348 | 2.993 | 1.911 | 4.027 | 7 |
| 32 | 5.057 | 4.003 | 4.616 | 3.951 | 3.185 | 4.302 | 5.114 | 2.219 | 3.432 | 2.056 | 10 |
| 33 | 2.826 | 1.100 | 1.681 | 1.883 | 2.931 | 0.752 | 2.042 | 2.156 | 2.223 | 3.917 | 6 |
| 34 | 3.631 | 1.586 | 2.894 | 0.495 | 3.454 | 1.949 | 3.305 | 2.136 | 3.040 | 4.180 | 4 |
| 35 | 3.372 | 2.338 | 1.754 | 3.189 | 3.329 | 0.818 | 0.630 | 2.758 | 1.846 | 3.944 | 7 |
| 36 | 3.383 | 1.980 | 1.906 | 2.551 | 3.397 | 0.521 | 1.320 | 2.578 | 2.105 | 4.077 | 6 |
| 37 | 4.330 | 2.345 | 3.747 | 0.399 | 4.169 | 2.797 | 4.132 | 2.764 | 3.856 | 4.793 | 4 |
| 38 | 1.946 | 0.607 | 1.960 | 2.110 | 2.259 | 2.237 | 3.369 | 2.190 | 2.886 | 3.851 | 2 |
| 39 | 2.693 | 2.320 | 2.304 | 3.543 | 0.697 | 2.739 | 3.321 | 1.411 | 1.743 | 1.590 | 5 |

【 図 2 4 】

| 文書No. | 文章 | 確率 | | | | 平均結果 | | | | 最も近い 分類番号 | 正答率 |
|-------|---|-------|-------|-------|-------|------|----|----|----|--------------|--------|
| | | 楽観 | 疑問 | 感嘆 | 苦情 | 楽観 | 疑問 | 感嘆 | 苦情 | | |
| 30 | コンフォートサポートソックスにクリーム色・無地し ーツにシャンパンは できませんか。 | 81.20 | 12.74 | 16.30 | 5.32 | 1 | 0 | 0 | 0 | 7 | 94.61% |
| 31 | ボディ美白ローションの替えを出して欲しいです。 | 90.82 | 0.60 | 7.69 | 15.58 | 1 | 0 | 0 | 0 | 4 | 94.61% |
| 32 | 前向適用サンスクリーンを買って使用していますが、つけたあとベ タつきがあり、オイルリムーブで油分を取ってからリターナティブシェ ンをつけています。これはオイルカットではないのですか？以後もこの方法で 使用してよろしいですか。 | 0.00 | 96.52 | 8.89 | 61.13 | 0 | 1 | 0 | 1 | 10 | 85.86% |
| 33 | (1) 偏科のソックスを使用していましたが、ロゴはきつすぎると思います。 20KMの夏ハイキングに使用したら、ぬいだ時の足の赤い處とただれ がしっかりつき、露白はスカートがはけませんでした。自宅用にも不向き。 (2) スポーツブラをもう一度検討して下さい。 | 47.83 | 1.76 | 25.89 | 50.03 | 0 | 0 | 0 | 1 | 6 | 49.84% |
| 34 | 4121 相模コーデン 1本のみ破損 (化粧箱裏が濡れてやぶれてい た) 賞状期限00.6.3 返品交換済 | 12.22 | 0.77 | 3.98 | 69.86 | 0 | 0 | 0 | 1 | 4 | 90.84% |
| 35 | クレンジング剤やサンスクリ ン剤、シャンプーはミニサイズがあると良い ですね。旅行の度に何個もボトルを削減して詰め替えるのは手間なの で……。 | 76.30 | 1.67 | 10.19 | 24.01 | 1 | 0 | 0 | 0 | 7 | 41.29% |
| 36 | プレストバやダー化粧直し用のケースがもう少し薄くスリムになるといいと思 います。他のファンデ ーションケースよりも厚く、小さなバッグの時に重いの です。 | 71.02 | 0.00 | 12.84 | 41.96 | 1 | 0 | 0 | 0 | 6 | 49.84% |
| 37 | グッドサポートソックス PB 一座の洗濯で変色 | 6.07 | 0.00 | 2.13 | 92.51 | 0 | 0 | 0 | 1 | 4 | 90.84% |
| 38 | 注文サンプルと間違っく入ってましたお詫します。足立 | 0.00 | 5.65 | 47.85 | 45.99 | 0 | 0 | 0 | 0 | 2 | 37.21% |
| 39 | UV CUT67をつけて毎年ゴルフもしても日焼けせず安心していましたが以 前の長いフェュープ入りの方が日焼けしなかったように思いますが？何時 も、UV、CUT57+UV CUT40をつけています。ゴルフ大好き60才台 の主婦には「美白BB」が最適ですか？ | 12.05 | 54.53 | 87.32 | 17.68 | 0 | 1 | 0 | 0 | 5 | 49.84% |